

TANULMÁNYOK

DR. JAKOBI ÁKOS

A grid: aggregált és dezaggregált rácsmodellek a területi egyenlőtlenségek vizsgálatában

A rácsmodellek iránti igény a társadalom területi kutatásában

Annak ellenére, hogy a raszteres adatfeldolgozásra a társadalmi-gazdasági vizsgálatokban eddig is bőven adódtak lehetőségek, még mindig viszonylag szűkös azon példák köre, amelyek képesek elrugaszkodni a társadalom- és gazdaságstatisztika hagyományos közgazdasági térfelosztáshoz illeszkedő poligonokat használó vektoros szemléletétől (lásd pl. Eurostat 2006). Pedig az effajta megközelítés igen jól támogatja az ún. módosítható területi egység problematika (MAUP = *modifiable areal unit problem*) egyik lehetséges, bár nem abszolút érvényű megoldását. A legtöbb társadalom- vagy gazdaságstatisztikai felületkartogram módszert alkalmazó térkép valamelyest torzítva közvetíti az információkat azáltal, hogy a nagyobb közigazgatási egységek poligonjai nagyobb foltokként, míg a kis területűek esetenként alig látható apró kitételekként látszanak (tipikusan ez a helyzet Magyarország települési adatait ábrázoló térképein is, ahol az alföldi, általánosságban nagyobb poligonok foltjai feltűnőbbek, mint a dunántúli települések kisebb egységei). Az ország területére helyezett egységes rácsméretű modellek azonban képesek eltüntetni az ország egyik és másik része között a térfelosztásból adódó részletezettségbeli különbségeket, a kérdés ezek után már csak a rácsméret optimális megválasztása marad.

A rácsmodellek legfontosabb előnye tehát az azonos cellaméret, ami az adatok összehasonlítása szempontjából kiemelt jelentőségű, ám a gridek használata mellett más érvek is felsorakoztathatók. A gridek területi beosztása például stabil az időben, nem érintik az adminisztratív változtatások. A gridekhez továbbá könnyen hozzákapcsolhatók más szakterületek vagy tudományterületek térbeli adatai (pl. klimatikus adatok), nincs szükség például adminisztratív egységekhez illeszkedő korrekciókra. A gridek celladatai könnyen aggregálhatók bármiféle hierarchikus rendszerbe, sőt a gridek segítségével könnyen összeállíthatók speciális területi aggregátumok is, melyek pusztán az adott vizsgálat szempontjából lehetnek fontosak (lásd pl. hegyvidéki területek).

A grid-szerű adattárolás, illetve adatközlés lehetőségeit természetesen a nemzetközi statisztika is felismerte. A statisztikai hivatalok egyike-másika rövidebb-hosszabb ideje közöl már gridformátumú adatokat vagy elemzéseket (lásd pl. SEDAC 2000), köztük ezt a lehetőséget újabban a KSH is tervezi. A gridadatok használatát az Európai Unió is szorgalmazza az európai téradat infrastruktúra (INSPIRE) irányelvéhez igazodóan kialakított rácsadat-rendszerekben, melyhez kapcsolódóan a KSH honlapja már magyar példákat is közzétett (lásd www.ksh.hu/inspire). Az Eurostat 2010-ben indult GEOSTAT programja a fentiekhez kapcsolódóan ugyancsak a gridadatok használatának elterjesztését szorgalmazza. Ennek keretében a tagállamok 2011-es népszámlálási adataiból egységes, nagy részletezettségű gridet alakítanak ki.

Bár a rácsmodelleket többségében pusztán a gridalapú adatközlés miatt hozták létre, egyes példák kifejezetten a griden értelmezhető elemzések miatt használnak rácsokat. Az amerikai Yale egyetem G-Econ projektje például a grid módszertanát hasznosítva alkotta meg a több évre is meghatározható bruttó cellánkénti termék (GCP, *gross cell product*) jelzőszámát (G-Econ 2011). Az indikátor a globális fókálózat által kijelölt fokcellákat felhasználva becsülte meg az egyes cellákba eső területeken előállított bruttó hazai termék nagyságát.

A gridek vagy rácsmodellek az egységes cellák használata révén nemcsak a térfelosztás homogenizálására alkalmasak, de raszterjellegű karakterisztikájuknak köszönhetően a raszterszerű elemzési technikáknak is teret engednek, annak ellenére vagy amellett, hogy mégis megtartják vektoros adattárolási formájukat. Ez a tulajdonság összességében számos praktikus lehetőséget kínál a társadalom és a gazdaság területi kutatói számára amellett, hogy új megvilágításba is képes helyezni a korábbi területi vizsgálati tapasztalatokat.

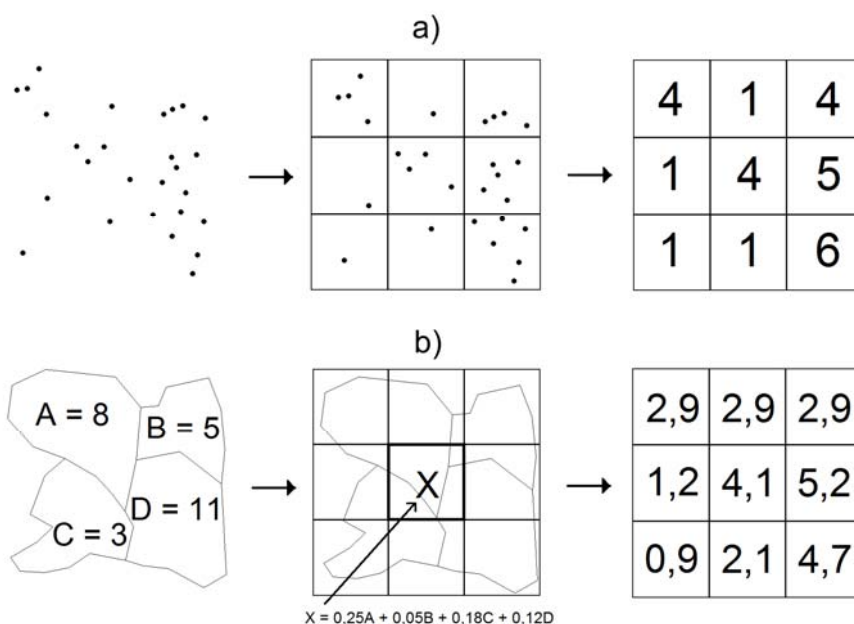
A rácsmodellek kialakításának lehetséges módjai

Társadalmi és gazdasági adatokból alapvetően két módon képezhetők rácsmodellbe szervezett raszter-szerű grid adatok. Az alulról építkező aggregáló jellegű (*bottom-up*) módszer a pontmintázatok elemzési elvét (Fotheringham–Brunsdon–Charlton 2000), valamint például a kvadrát analízis közelítésmódját (Thomas 1977) kihasználva, a pontthalmazokra fektetett négyzethálók segítségével ad gridszerű eredményeket. Ez az egyszerűbb és egyben tisztábbnak is tekinthető eljárás, hiszen a pontszerű adatok aggregációja során nem tapasztalható információtorzulás.

A raszterszerű gridadatok kialakításának másik módja a már meglévő vonalas vagy felület adatok dezaggregálásának (*top-down*) módszerével dolgozik. Ez esetben a meglévő egységeket bontjuk szét a rácshálózathoz illeszkedő részelemekre, majd ezeket a részelemeket szervezzük újra a gridnek megfelelő aggregációban (lásd például Funk–Stiefer–Chinkin 2001, vagy Nordhaus et al. 2006). Bár ez az eljárás számításgényesebb, illetve az adatok dezaggregálásával létrehozott új adatértékek csak becslésként értelmezhetők, mégis haszonnal alkalmazhatók akkor, ha a kiindulási adatok eleve csak poligon-, azaz nem pontszerű formában állnak rendelkezésre. Ez az eljárás gyakoribbnak tűnik a közigazgatási egységekhez illeszkedő statisztikai adatok utólagos rácshálóba szervezésekor.

1. ábra

A rácsmodellek kialakításának aggregáló bottom-up (a) és dezaggregáló top-down (b) módszere



Forrás: saját szerkesztés.

Az aggregáló jellegű *bottom-up* módszer (1. ábra a rész) kizárólag pontszerű (vagy pontokként értelmezhető) adatok feldolgozásánál hasznosítható. A térinformatikai rendszerekben a pontrétegekre helyezett eredendően kvantitatív attribútumokkal még nem rendelkező rácsáló egyes celláiba területi aggregáló metszéssel (spatial join/sum) számítható ki az oda eső pontok darabszáma. Ez a cellánként összesített érték értelmezhető a grid cellánkénti súlyának is. A rácsáló celláinak további kvantitatív attribútumai hasonlóképpen lokális összesítéssel határozhatók meg, így például az adott cella összesített népességszáma az oda eső pontok népességszámának összegeként adódik. Az összesített attribútumérték (pl. össznépességszám) és az adott cellába eső pontok összesített számának (a cellánkénti súlynak) hányadosaként meghatározható a cella átlagos attribútumértéke is (számos vizsgálat inkább ezen szám adatokra épül).

A területi adatok részarányos szétbontására épülő dezaggregáló *top-down* módszer (1. ábra b rész) poligonokban (esetleg vonalakban) értelmezett adatokból indul ki. Az eljárás területileg átlapolat adatszámítással dolgozik, melynek során az eredeti poligonalakzathoz nem illeszkedő, csak részben átfedő új gridpoligonok attribútum-értékét az átfedő poligonrészek arányában határozzuk meg. Az új célpoligon attribútumértéke az eredeti poligonok annyiad részeinek összegzett értékeként adódik, amekkora részben az új és az eredeti poligonok átfednek. Ha például az X cellába eső összesített népességszámot szeretnénk kiszámítani, és az adott cella az eredeti A terület 25, a B terület 5, a C terület 18 és

D terület 12%-át átfedi, akkor X értéke az eredeti területek népességszámának (vagy hasonlóképpen más egyéb attribútumának) ismeretében az alábbi módon határozható meg:

$$X = 0,25A + 0,05B + 0,18C + 0,12D$$

Technikai értelemben a legtöbb térinformatikai szoftver több lépésben képes a fenti műveletek elvégzésére. Egy lehetséges megoldás, ha először az eredeti poligonokat a rácsháló beosztásához illeszkedően, azaz a rácsháló rétegét felhasználva darabokra metsszük (intersect), majd az egyes újonnan létrehozott daraboknak az eredeti „anyapoligonhoz” viszonyított területarányát számítjuk ki egy új adatmező formájában. A harmadik lépésben a vizsgált attribútumunk (pl. népességszám) „anyapoligontól” öröklött értékét az előző lépésben meghatározott területarányának megfelelő szorzótényezővel korrigáljuk, majd a negyedik lépésben a poligondarabokat a rácsháló szerinti beosztást követve cellákba aggregáljuk (dissolve), aminek az eredményeként nemcsak a poligonok geometriája, de a hozzájuk tartozó korrigált attribútum adatok is megfelelő módon összegződnek.

A fent ismertetett *top-down* módszer az aggregáló *bottom-up* módszerrel ellentétben nem tökéletesen pontos, csak becslő jellegű, mivel azt feltételezi, hogy a dezaggregáció alapját képező számszerű jelenség a szétbontásra kerülő poligonokban egyenletesen oszlik el. Nyilvánvaló módon kevésbé hasznosítható ez a technika akkor, ha a vizsgált jelenség szempontjából fontos a poligonon belüli tényleges elrendeződés.

A két alapmódszeren kívül persze egyéb alternatívák is lehetségesek. Az Eurostat népességi gridmetodológiája például azokban az országokban, ahol a népességi adatok nem érhetők el georeferált címpontos formában, speciális dezaggregációs módszertant javasol egy közbülső lépés beszúrásával (Eurostat 2006). Az eljárás abból a feltételezésből indul ki, mely szerint a lakónépesség lényegében azokon a területeken tömörül, ahol az épített környezet sűrűsége, jellege is ezt visszatükrözi. Az ilyen karakterisztikájú területek a Corine távérzékeléses felszínborítottsági adatbázisból egyszerűen kiválaszthatók és leszűrhetők, s így létrehozhatók az adott közigazgatási körzetben valóban lakott területek övezetei is. Ezt követően a közigazgatási egység eredeti népességszámát a modell már csak a lakott területeken veszi figyelembe, ráadásul az eljárás azt is feltételezi, hogy a beépítési sűrűség és a népsűrűség között lineáris korrelációs kapcsolat van, következésképpen nagyobb lakosságszámot becsül a sűrűbb beépítésű, míg kisebbet a ritkább beépítésű zónákba.

Módszertanilag mindezek után (vagy éppen mindezek előtt) már csak az döntendő el, hogy a rácsmodellünkben optimálisan mekkora cellaméretet válasszunk. Univerzális megoldás vagy egyenlet erre a problémára nem fogalmazható meg, az azonban egyértelműnek látszik, hogy túlzottan nagy cellaméret használatakor nagyobb az esély a vizsgált jelenség térbeli varianciájának elmosódására, míg nagyon kis cellaméret alkalmazásakor (főleg a *bottom-up* módszer esetében) a lokálisan egyedi sajátosságok befolyásoló szerepe vagy az adathiányos cellák előfordulási esélye növekszik meg. Egyes térinformatikai szoftverek ugyan javaslatot tesznek az optimális cellaméret meghatározására (az ArcGIS-ben a vizsgált teljes ponthalmazt vagy területet lefedő téglalap rövidebb oldalának 250-ed részeként adódik a javasolt cellaméret), de ezek az automatizmusok nem igazán veszik figyelembe a vizsgált alapegységek számát, nagyságát vagy területi eloszlását. A rácsméret helyes meghatározásakor javasolt tehát a fent említett paraméterek előzetes vizsgálata (lásd Hengl 2006) (a témához kapcsolódó parametrizálható gridkalkulátor elérhető az

alábbi weboldalon: http://spatial-analyst.net/wiki/index.php?title=Grid_size_calculator). A rács-méret kiválasztásakor persze egyéb szubjektív vagy objektív szempontok is döntőek lehetnek. Túl kis cellaméret alkalmazásánál például gyakori lehet az egyedi adat-előfordulások szerepe (főleg a *bottom-up* módszer esetében), így a hivatalos adatkezelés/adatkiadás során az adatvédelmi szempontokat figyelembe vevő kényszerű adattorzításra is sor kerülhet.

Pszedoraszter térstatisztikai műveletek

A grid típusú adatkezelés lényegében két módszertani célt szolgál ki. Az első a vizualizációhoz kapcsolódik, amikor a gridszerű adatok alkalmazása csak a rácshálóban aggregált vagy dezaggregált adatok megjelenítésére fókuszál. E viszonylag gyakori és valamelyest hagyományosabbnak is tekinthető eljárás célja pusztán az adminisztratív beosztásoktól független adatmegjelenítés. A gridmodelleket használó szakmai kör legnagyobbbrészt megáll a rácshálók efféle hasznosításánál, pedig nem kevés lehetőséget kínál a gridek felhasználásának másik módja, a pszedoraszter térstatisztikai adatelemzés.

A raszterszerű rácsmodellekben való gondolkodás iránti igény maga után vonta azt is, hogy a raszterelemzés klasszikus módszertanát is használni lehessen ezeken az adatokon. A vektoros környezetben létrehozott és tárolt rácsháló és a raszter-analitikai módszertan összekapcsolása lényegében egy ál-raszteres vagy pszedoraszter metodika kialakításához vezetett, hiszen továbbra is vektoros adatokkal dolgozunk, de mégis a raszterelemzés eszközeit használjuk a vizsgálatokban.

A pszedoraszter elemzési eljárások logikája igen egyszerű. Elsőként a kiindulási forrásként használt társadalmi-gazdasági tartalmakból a fent említett *bottom-up* vagy *top-down* módszerrel vektoros rácsmodelleket szükséges szerkeszteni. A kialakított grid cellaértékei, mint attribútumok megtartása mellett ezt követően a gridegységeket raszterekké lehet konvertálni, vagy legalább is raszterekként lehet értelmezni. A kapott eredmények pszedoraszterekként funkcionálnak, melyeken a legtöbb raszteres művelet értelmes módon kivitelezhető.

A hagyományos raszter-elemzési módszertan számos eljárást ismer, melyek közül több is jól használható a társadalmi-gazdasági differenciák rácshálóban értelmezett területi vizsgálatainál. Jól használható eljárásokat kínálnak például a területi szűrők vagy szomszédsági operátorok.

A területi szűrők (filterek) eljárásainak középpontjában a rasztertopológia áll, más szóval az a tény, hogy a raszteres rendszerben a képpontok (pixelek) determinisztikus szomszédsági rendben, azaz általában – de nem kizárólagosan – négyzethálós formában helyezkednek el egymás mellett, s így minden cella szomszédsági rendszere könnyedén definiálható. A szűrők lényegében raszteres szomszédsági operátorok, amelyek az alkalmazott algoritmusok alapján a szomszédsági adatok ismeretében alakítják át az adott lokációban található raszter értékét. A raszterelemzés ún. fokális operátorainak használata során az outputértékek a vizsgált cellahalmaz fókuszában realizálódnak. Más szóval, ha például egy 3x3 cellás ablakkal futtatjuk a fokális átlag operátort, akkor az átlagszámítás eredményét az ablak középpontjaként szolgáló cellába kapjuk. A fokális operátorok új értékei tehát a szomszédos entitások eredeti értékein alapulnak. A raszteres adatbázisok helyi szomszédsági vizsgálataihhoz alkalmazható függvények köre igen változatos (lásd

Tomlin 1990), ilyen például a fokális átlag, a fokális medián vagy a fokális diverzitás. Míg előbbi egy egyszerű vagy súlyozott átlagot ad eredményül a centrális raszterben, addig a fokális medián a vizsgált 9 érték mediánját számítja ki. Ha a raszterben szereplő értékek kategóriákat jelölnek (például földhasználati vagy beépítési övezeteket), akkor a fokális diverzitás (vagy fokális változatosság, más néven NDC = number of different classes, különböző osztályok száma) a különböző értékek számát fogja megszámlálni a 3x3-as mátrixban (a szűrő mérete természetesen más is lehet, beszélhetünk pl. 5x5-ös, 7x7-es stb. szűrőkről is).

A különféle statisztikai funkciók használata a területi szűrésben más és más eredményeket adhat. A felhasználási lehetőségek és az alkalmazható eljárások köre tehát igen változatos. A teljesség igénye nélkül többek között az alábbi eljárások említhetők:

- a cellák értékének átlaga,
- a leggyakrabban előforduló érték,
- a legnagyobb érték,
- a cellák értékének mediánja,
- a legkisebb érték,
- a legritkábban előforduló érték,
- a legnagyobb és a legkisebb értékek különbsége,
- a cellák értékének szórása,
- a cellák értékének összege,
- a cellák között az egyedi előfordulások száma.

Hagyományos értelemben a raszteres szűrés azokat az eljárásokat foglalja magában, amelyekkel az eredetileg nyert értékekről az esetleg fellelhető (többnyire zavaró jellegű) hatásokat leválasztják, vagy más esetben külön kiemelik (Detrekői – Szabó 2002). A területi szűrésnek jellegük szerint így lehetnek aluláteresztők (*low pass filters*), melyek célja a zavaró jellegű hatások kiküszöbölése, vagy feluláteresztők (*high pass filters*), ahol a cél a zavaró jellegű hatások kiemelése. Ezek az eljárások a társadalmi-gazdasági adatokból képzett pseudoraszter rácsmodelleknél is hasonlóképp használhatók.

Az aluláteresztő simító szűrő az extrém adatértékek hatásait mérsékli (lényegében az izolált értékeket távolítja el), többnyire a szomszédok értékeinek egyszerű átlagolásával. A szűrő által alkalmazott konvolúciós maszk a 3x3-as ablak összes cellájára az alábbi egyenlő súlyozást alkalmazza, ahol w a középponti cellába kapott súlyérték:

$$w = \frac{1}{9} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

A modellben minden egyes cella értéke összességében 9 cella adatából származik, kivéve a rácsháló szélein, ahol vagy kisebb inputelemszámú a szűrő, vagy egyáltalán nem végez számítást a modell. A gyakorlatban ha a szélső cellákat nem vesszük figyelembe, akkor az output cellahalmaz kisebb lesz az eredetinel.

A módusz- (fokális többség-) és a mediánszűrő ugyancsak simító jellegű. A mediánszűrő előnye, hogy az átlaggal ellentétben nem érzékeny az outlier-ekre, a módusz-szűrő pedig nominális cellaadatokon is alkalmazható.

Alternatív megoldást kínál a gridállomány simítására a kernelszűrő, amely aszerint súlyozza a cellákat, hogy milyen közel találhatók azok az ablak közepéhez. Erre példák az alábbiak:

$$w = \frac{1}{10} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

$$w = \frac{1}{16} \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}$$

A simító hatású szűrőkkel ellentétben a felüláteresztő szűrők épp ellenkező eredményt adnak: kiemelik az éleket, növelik a magas intenzitású értékeket és csökkentik az alacsony intenzitásúakat. Az élkiemelő szűrők azonban a zajt is növelik. Ezek a szűrők kihangsúlyozzák azokat a területeket, melyeknek szomszédságán belül szemmel láthatóan különbözőek a cellaértékek. Az élkiemelő operátorok a szomszédos cellák szignifikáns változásainak azonosítására törekednek, a hirtelen változású területeket keresik. A szűrők célja tehát meghatározni azokat a határvonalakat, ahol a változások aránya a legmagasabb, de további cél lehet az egynemű területek kihangsúlyozása is. Míg az aluláteresztő szűrők ablakértékösszege 1, addig az éldetektáló (felüláteresztő) szűrőknél ez az összeg 0, ezért azokat a cellákat, melyeket azonos értékű cellák vesznek körül zérussá változtatja. A Laplace-féle élkiemelő szűrő az alábbi formában határozható meg:

$$w = \begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

A pszeudoraszter alapú topológikus (tehát a cellák helyzeti szomszédságát kihasználó) területi szűrők meglehetősen nagy szabadságot nyújtanak a felhasználóknak. A szűrők maguk rendkívül sokféleképpen definiálhatók. Az ablakméret tetszőleges $N \times N$ formában meghatározható, a belső cellasúlyok változatosan alakíthatók (általában egész számok, de lehetnek törtek is). Általánosságban egy $N=3$ szélességű mátrixban a területi szűrő az alábbi tipikus, de nem kizárólagos formát öltheti:

$$\begin{pmatrix} c & a & c \\ a & b & a \\ c & a & c \end{pmatrix} \text{ vagy } \begin{pmatrix} d & e & f \\ g & h & i \\ j & k & l \end{pmatrix}$$

ahol a , b és c tetszőlegesen megválasztható, akár egyenlő is lehet (pl. a simító szűrő esetében), vagy csak a és c egyenlő (pl. kernelszűrőknél), vagy lehet akár d -től l -ig mindegyik érték különböző. Még általánosabban fogalmazva ha Z_i egy $m=N \times N$ mátrixban (ablakban) található együtthetők halmaza és P_i az ablakba eső kiindulási cellaértékek halmaza, akkor adott szűrt pszeudoraszter-cella (R) értéke

$$R = \sum_{i=1}^m Z_i P_i \bigg/ \sum_{i=1}^m Z_i + B$$

ahol B értéke, mely leggyakrabban 0, egy módosító tényező, ami növeli vagy csökkenti a filter értékét egy fix mennyiséggel (Fotheringham–Brunsdon–Charlton 2002). A fentebb említett logika csak a lineáris szűrőkre vonatkozhat, a nem lineáris változatok (például amelyek a minimumot, a maximumot, a szórást vagy a különböző osztályok számát adják eredményül) az ablakba eső halmazon saját függvényeiket, algoritmusait futtatják le.

Egészen más elemzési lehetőségeket kínál a rasztermatematika vagy raszteres térképi algebra lokális változata. A raszteres térképi algebra lokális funkciói az egyes egyedi cellák adataival dolgoznak és csakis annak az adatait veszik figyelembe. A lokális térképi algebrai funkciók lehetnek aritmetikaiak, trigonometrikusak, exponenciálisak, logaritmusok, statisztikusak vagy logikaiak. Triviális példa, ha egy input réteget használunk és minden egyes cellát egy konstans értékkel szorzunk meg. Ennél izgalmasabb persze az, ha bemenetként már eleve több pszeudoraszter-állományt használunk és az egyik pszeudoraszter-réteg például a másik szorozójaként kap szerepet. A pszeudoraszteres lokális térképi algebrai műveletek a legkülönbözőbb változatúak lehetnek. Alkalmazásuknak lényegében csak a kutatói, felhasználói fantázia szab határt.

A gridállományokon a fentiek továbbvezetéseként komplex statisztikai műveletek is lefuttathatók, közülük a pszeudoraszter-állományok statisztikai összehasonlításának módszertanát célszerű példaként kiemelni. Egyértelmű, hogy két tökéletesen átfedő gridállomány azonos helyzetű cellái könnyű szerrel párba állíthatók (a tökéletes átfedés nemcsak méretbeli egyezést, de vetületi egyezést is jelent), és mivel ezen cellákhoz számszerű attribútum adatok tartoznak, így a páronkénti számadatok is összehasonlíthatók, azaz a gridből derivált adatsorok korrelálthatók. A pszeudoraszter-korreláció tehát azt vizsgálja, hogy az egyik gridréteg magas attribútumértékeinek megfelelő helyzetben a másik gridrétegben vajon magas vagy alacsony értékek szerepelnek-e. A pszeudoraszter-korreláció globális változata mindezt egy jól ismert korrelációs együtthatóval fejezi ki. A területi hatások hangsúlyosabb figyelembevételét is lehetővé tevő szomszédsági változat a fentivel ellentétben jóval erősebb térinformatikai igényeket is mutat (Berry 2013). A gridállományok helyi szomszédsági korrelációja során egy adott cella előre definiált sugarú környezetében lévő cellahalmaza és a másik grid ugyanazon helyzetű cellahalmazának adatértékeit korreláltatjuk. A mozgóablakos eljárás minden különálló helyzetű cella esetében egyedi korrelációs együtthatót ad eredményül. Ezzel az eljárással nem csak az vizsgálható, hogy a két kiindulási gridállomány összességében mennyire hasonlít egymásra, de az is, hogy az állományok mely földrajzi helyzetű pontjaiban mutatkozik jellemzően nagyfokú párhuzam, illetve hol tapasztalható gyenge vagy éppen fordított irányú hasonlóság. A rácsmodellekre értelmezett pszeudoraszter statisztikai műveletek a fent említett eljárások koncepciója szerint lényegében bármilyen statisztikai módszertan esetében végrehajthatók. Hasonló módon definiálhatók például grideket használó két vagy annál több változós regressziós modellek, vagy azok helyi szomszédsági változataként a rácsmodelleken futtatott földrajzilag súlyozott regressziós (GWR) számítások stb.

A rácsmodellek alkalmazási példái a városi és falusi területek vizsgálatában

Még kihasználatlanok, de már nem teljesen új keletűek a rácsmodelleket alkalmazó módszertanok a rurális és urbánus területek lehatárolásában, valamint a falvak és a városok közötti különbségek vizsgálatában. Az Eurostat számos módszertant használ az urbánus-rurális körzetek meghatározására, közülük az egyik legújabb tipológia 1x1 km-es gridcellákkal operál (Eurostat 2013). Ez a módszertan az urbánus területek azonosítását több lépésben oldja meg. Elsőként leválogatja a 300 fő/km² feletti népsűrűségű gridcellákat, majd ezt követően (az átlókat is figyelembe vevő) szomszédsági elv alapján csoportosítja az egységeket. Ha egy vizsgált cella értéke meghaladja a népsűrűségi küszöböt, akkor az összes olyan szomszédos cellával egyesítésre kerül, amely szintén 300 fő/km² feletti népsűrűségű. Végül urbánus területeknek azokat a körzeteket tekinti a modell, amelyekben a csoportosított gridcellák összesített lakosság száma meghaladja az 5000 főt. A metodika így egyszerre veszi figyelembe a népsűrűség és a lakosság abszolút számának nagyságát (a rurális területek értelemszerűen a fentiekén kívüli gridcellák lesznek). Az Eurostat továbbvezetett módszertana emellett elkülöníti a központi városi körzeteket is, ahol a gridcellák népsűrűsége 1500 fő/km² feletti és a csoportosított szomszédos cellák össznépessége meghaladja az 50 000 főt.

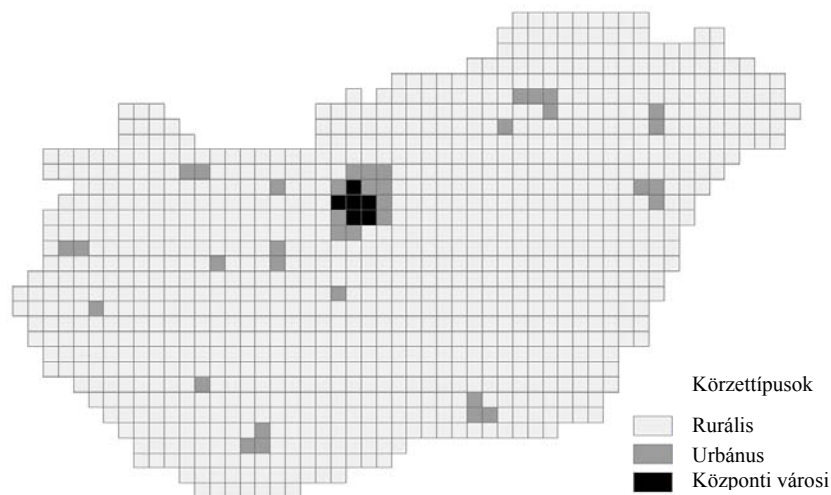
Az urbánus és rurális területek fenti elvek szerinti elkülönítése alternatív módon kiegészíthető vagy helyettesíthető a közigazgatási jogállásból származó információk modellbe integrálásával. A falusi-városi népességarány eddigi legalacsonyabb értelmezhető területi szintje a járás (kistérség) volt, annál kisebb léptékben, például települések szintjén ez a ráta nem meghatározható, hiszen egy település jogállás szerint vagy város, vagy község. Rácsháló-szerű aggregációban ugyanakkor vizsgálható ez a kérdés, ha a *bottom-up* módszerrel kialakított népességi aggregációkban attribútumként az eredeti települési jogállás státuszát is összesítjük. Modellünket az ESRI-BM 2012-es címpontos népességi adatait felhasználva alakítottuk ki, ahol minden lakossági pontadathoz hozzárendeltük az adott pont helyzetének megfelelő település közigazgatási jogállás szerinti attribútumát (a 2012-es állapotok szerint). A pontadatokat ezt követően a *bottom-up* módszernek megfelelően 10x10 km-es rácshálóba aggregáltuk, aminek eredményeként minden cellában meghatározhatóvá vált a teljes lakosság szám, valamint a falvakban illetve a városokban lakók összlétszáma, melyből már könnyedén kiszámítható volt a falvakban vagy éppen városokban lakó népesség lokális aránya is. A kapott eredménytérkép (2. ábra) ennek alapján szemlélteti a város-falu különbségeket a népesség területi megoszlásában.

2. ábra

A falvakban lakók arányának 10x10 km-es rácshálóban értelmezett értékei, 2012*Forrás: saját szerkesztés.*

A térkép előnye tehát, hogy képes a rátát alacsony területi szintre vagy legalábbis részletes területi bontásban szemléltetni, hátránya viszont, hogy a nagy külterületű, de viszonylag kis belterületi lakott résszel rendelkező (főleg alföldi) városok esetében városiasnak tekinti azokat a gridcellákat is, ahol valójában nem jellemző az urbánus környezet. Megoldásként az Eurostat fent említett módszertanához nagyban hasonló eljárást választottunk, azaz elsőként meghatároztuk azokat a cellákat, ahol a népsűrűség nagysága meghaladta a 300 fő/km²-es (illetve az 1500 fő/km²-es) értéket, illetve egy 3x3-as ablakban a fokális összeg módszerével minden cellára meghatároztuk a helyi körzetben összesített lakosságszámot. Mivel az Eurostat módszertanától eltérően itt nagyobb, 10x10 km-es rácsokkal dolgoztunk, urbánus területeknek azok a cellák adódtak, amelyekben a küszöbszintet meghaladó népsűrűség mellett a 3x3-as ablakban aggregált népességszám meghaladta az 50 000 főt (központi városi körzetek esetében az 500 000 főt). A modell alapján elkülönített rurális, urbánus és központi városi cellák elhelyezkedését a 3. ábra szemlélteti.

3. ábra

Rurális, urbánus és központi városi körzetek 10x10 km-es rácshálóban, 2012*Forrás: saját szerkesztés.*

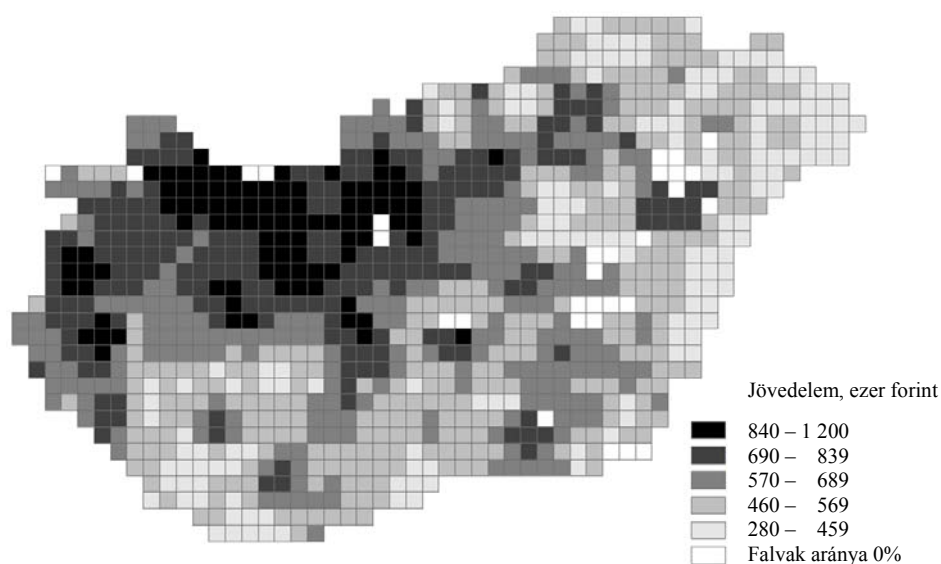
Gridmodellek a jövedelemszint területi differenciáltságának vizsgálatában

Az előző fejezet város-falu tematikájához kapcsolódva a következőkben a gridek körzet-lehatároláson túlmenő alkalmazási lehetőségei közül mutatunk be néhányat. A népesség-szám és a népsűrűség klasszikus rácsmodelljein túl a gridek jól alkalmazhatók egyedi csoportok országos területi karakterisztikáinak vizsgálatára, ami különösen hasznosnak tűnhet például azokban a jövedelmi egyenlőtlenségi elemzésekben, amelyekben nem a teljes sokaság, csak valamely részhalmaz területi differenciáltságát szeretnénk egyöntetű keretben vizsgálni. A következő rácsmodellek a falvak és a városok jövedelmi különbségeinek magyarázatához szolgálnak adalékul.

Az egy lakosra jutó adóalapot képező jövedelem területi különbségeit a hazai szakirodalom viszonylag alaposan vizsgálta (lásd pl. Németh–Kiss 2007, Péntes 2012), a falvak és a városok közötti jövedelemkülönbségek feltárása ennek viszont csak kisebb halmazában (Dusek 2004), vagy csak közvetett formában jelenik meg (Nemes Nagy–Jakobi–Németh 2001). A fenti elemzések tapasztalatai alapján a hazai településállományban nemcsak a falvak és a városok között, de a falvak és a városok önálló halmazain belül is komoly területi egyenlőtlenségek feltételezhetők. Ez tükröződik vissza a 4. és az 5. ábrán, amelyek kizárólag a falvakban élő lakosság, illetve csak a városlakó népesség vonatkozásában jelenítik meg a jövedelmi különbségeket. Modelljeink mindkét esetben a *top-down* metodikát követték, mivel a jövedelmi alapadatok kizárólag települési bontásban álltak rendelkezésünkre (TeIR NAV forrásból), így a *bottom-up* módszerhez szükséges egyéni szintű jövedelmek, illetve a településen belüli egyéni jövedelem-eloszlás nem volt meghatározható.

4. ábra

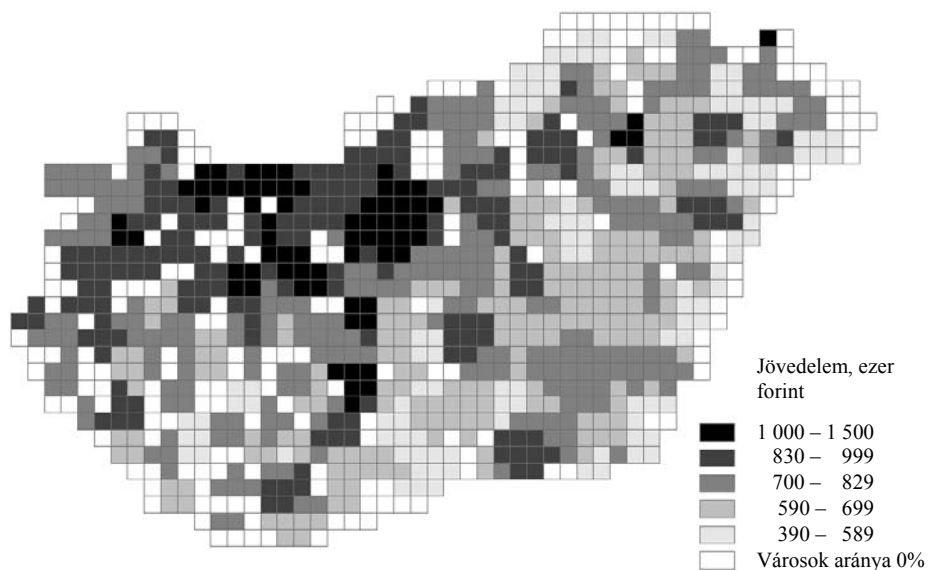
A falvakban élők egy lakosra jutó adóalapot képező jövedelme 10x10 km-es rácshálóban, 2012



Forrás: saját szerkesztés.

5. ábra

A városokban élők egy lakosra jutó adóalapot képező jövedelme 10x10 km-es rácshálóban, 2012

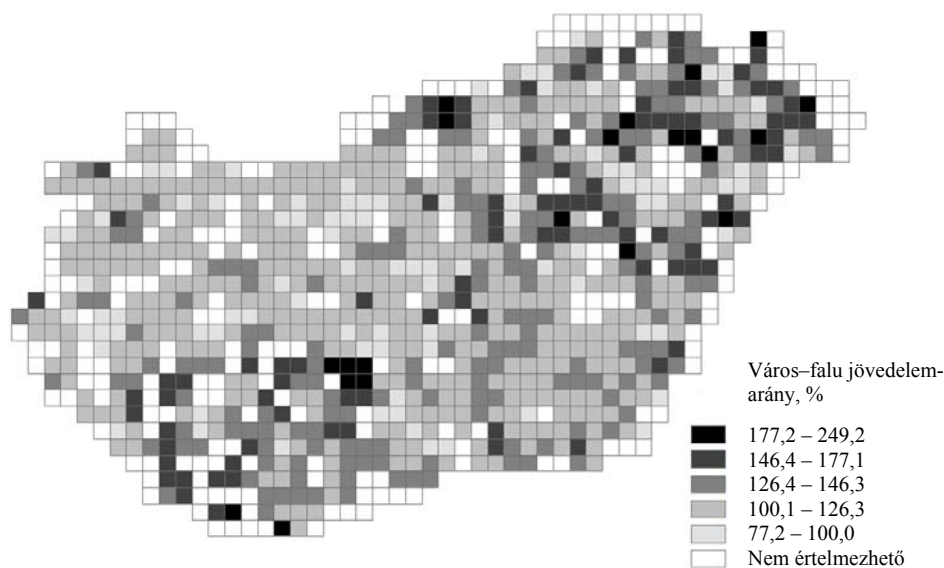


Forrás: saját szerkesztés.

A 4. ábra tehát a városok figyelembevétele nélkül mutatja a lakossági jövedelmek hazai területi különbségeit. A homogén méretű téregységekben jól megfigyelhetők az ország egyes részei közötti általánosan jellemző különbségek, például a Dunántúl északi és déli része közötti eltérések. Nyilvánvaló, hogy azokban a cellákban, ahol egyáltalán nem található falvak (pontosabban ahol a falvak területaránya 0%), nem lehet a falvakban élők jövedelemszintjét sem értelmezni, ámbar viszonylag kevés az efféle cellák száma. A városi adatok nélkül kirajzolódó kép közvetett módon ugyanakkor mégis visszatükrözi a jelentősebb városok elhelyezkedését, azaz általánosságban megfigyelhető, hogy a városi agglomerációk falvainak cellái a környezetükben található falvak celláinál valamelyest magasabb jövedelemadatokkal rendelkeznek (a laikus például Debrecen, Szeged, Pécs vagy éppen Zalaegerszeg stb. környékét is be tudja azonosítani). A kizárólag a városi területeket lefedő rácsmódel térképe már több adathiányos cellát is mutat (5. ábra), ennek ellenére ez a modell is alkalmas az adott halmazon belüli általános területi különbségek feltárására, szemléltetésére, ezúttal a városok esetében.

6. ábra

A városlakók egy főre jutó jövedelme a falvakban lakók egy főre jutó jövedelmének százalékában 10x10 km-es rácshálóban, 2012



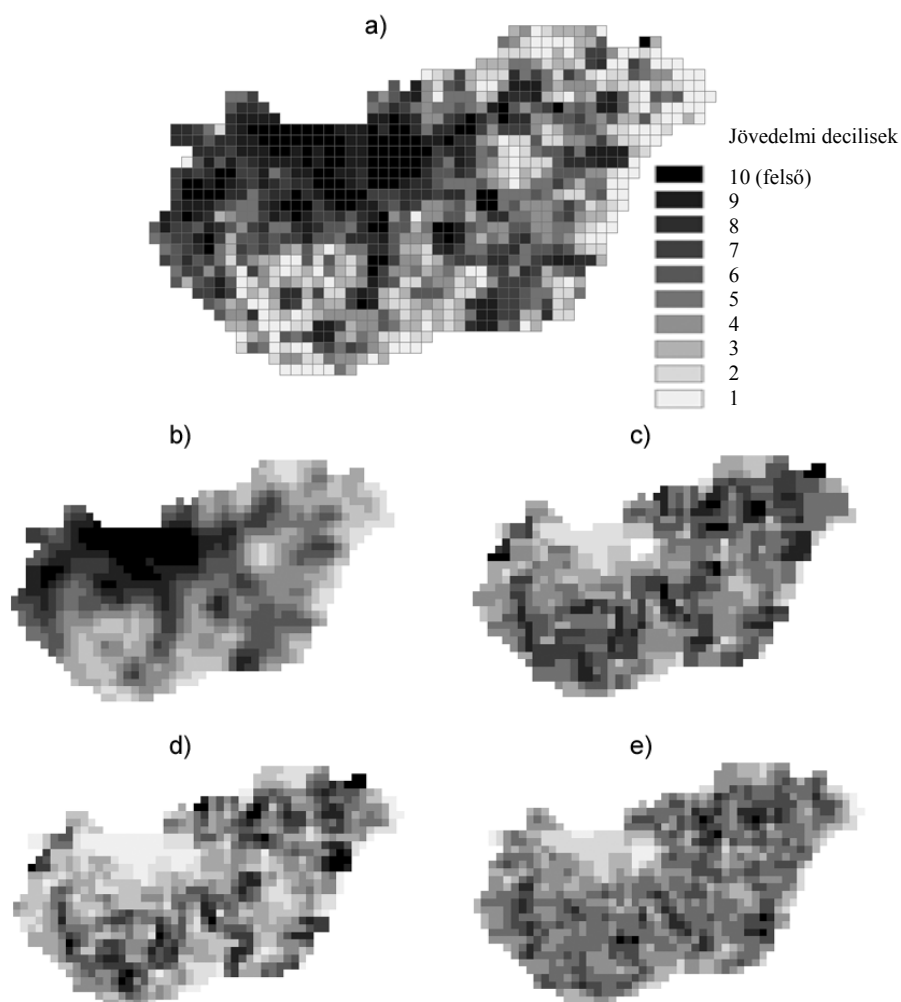
Forrás: saját szerkesztés.

A két fent említett rácsmódel – nem meglepő módon – hasonlóságokat is mutat, ami alapján feltételezhető, hogy egy település esetében inkább a földrajzi elhelyezkedés, s csak másodsorban a települési jogállás az, ami a jövedelemszintet befolyásolja. Mindez egyszerűen vizsgálható, ha a két rácsmódel a pszudoraszter metodika szerint összevetjük egymással. A 6. ábra ennek megfelelően a falvakra, illetve a városokra meghatározott egy főre jutó jövedelmi értékek cellánkénti relatív különbségét szemlélteti, más szóval az adott cellán belül értelmezett városi jövedelem nagyságát a cellán belül definiált falusi

jövedelem százalékában jeleníti meg. Az eredmények szerint a cellák többségében a városok jövedelme meghaladja a falvakban élők jövedelmét, de a számításba vehető cellák 11%-ánál a falusi jövedelem múlja felül a városit (ezek az esetek az országban szórtn fordulnak elő). A magas városi és alacsony falusi jövedelem leginkább a Dunántúl déli részén, valamint Észak- és Északkelet-Magyarország egyes celláiban jellemző. A legnagyobb különbség Paks és környékének cellái esetében mérhető. Az ábrán feltűnő továbbá, hogy a Dunántúl északi fele általánosságban kiegyenlítettebb, avagy itt mérsékeltebbek a városok és a falvak közötti különbségek.

7. ábra

Az egy lakosra jutó jövedelem 10x10 km-es rácshálóban meghatározott decilisenkénti megoszlása és továbbvezetett térképei, 2012



Forrás: saját szerkesztés.

A gridmodellek nominális, valamint ordinális adatok területi előfordulásainak vizsgálatában is sikerrel hasznosíthatók. Klasszikus téma például a jövedelmi tizedek szerinti csoportok területi vizsgálata (lásd Pénzes 2011), amely ordinális kategóriák térbeli eloszlása a gridmodellek segítségével is sokszínűen elemezhető. A következő példa néhány fokális operátor felhasználásával kialakított eredményt jelenít meg. A kiindulásként használt jövedelmidecilis-eloszlás ordinális kategóriáiból (7. ábra a/ résztérkép) az alkalmazott 3x3 cellás, azaz 30x30 km-es fokális modellek segítségével meg lehetett határozni a jövedelemszint fokális átlagértékét (7. ábra b/ résztérkép), terjedelmét, azaz a maximális és a minimális kategóriaérték különbségét (7. ábra c/ résztérkép), a helyi egyenlőtlenségeket kifejező szórását (7. ábra d/ résztérkép), vagy a jövedelmidecilis-kategóriák helyi változatosságát (7. ábra e/ résztérkép).

Míg az eredeti jövedelmi decilisek szerinti csoportokat megjelenítő ábra (7. ábra a/ résztérkép) a jövedelmi egyenlőtlenségek klasszikus hazai mintázatát sejteti, addig a modellből továbbvezetett térképek kevésbé megszokott térbeli vonásokat is feltárnak. A fokális átlag eredményeket megjelenítő térkép (7. ábra b/ résztérkép) a területi mozgóátlag modellek (Dusek 2004) eredményeihez hasonlatosan a jövedelemeloszlás generalizált ábráját mutatja például a fővárosi agglomeráció és a Dunántúl északi részének egyöntetűen magasabb jövedelmi kategóriáival. Ehhez képest a 3x3-as ablakban értelmezett fokális terjedelem eredménytérképe (7. ábra c/ résztérkép) határozott eltéréseket mutat. A jövedelemdecilis szerinti kategóriák maximális és minimális értéke az ország több területén is eléri a legnagyobb (9-es vagy esetleg 8-as) értéket, ami arra utal, hogy nem kevés azon 3x3-as méretű körzetek száma, ahol a legalsó és a legfelső decilisbe tartozó cella is előfordul. A modell felüláteresztő tulajdonságú és kvázi élkiemelő jellegű, aminek eredményeként láthatóvá válnak a nagy és a kis jövedelmű területek határvonalai (például Bács-Kiskun megyében). A fokális szórás módszere lényegében hasonló karakterisztikájú eredményeket kínál (7. ábra d/ résztérkép). A modell itt továbbá a területi egyenlőtlenségek lokális (3x3-as cellában értelmezett) indikátorának is tekinthető, így ahol az ez alapján kirajzolódó kép alacsony szórást jelez, ott a jövedelemkategóriák szerinti területi különbségek is mérsékeltek (jellemzően ezt tapasztalhatjuk a Dunántúl északi részén és a Budapest körüli cellákban), míg ahol a fokális szórás értéke viszonylag magas, ott a helyi környezetben mért jövedelemegyenlőtlenség foka is nagy lehet. A fokális változatosság modellje (7. ábra e/ résztérkép) végül azt jeleníti meg, hogy a vizsgált 3x3-as ablakba eső 9 cellában összességében hányféle különböző jövedelmi kategória fordul elő. A modell ismételten leginkább a főváros környékén és például Komárom-Esztergom megyében jelzett kisebb helyi változatosságot, míg például Heves és Borsod-Abaúj-Zemplén több körzetében nem egy esetben 9-ből 8 cella különböző deciliskategóriába esett.

Összegzés

A térinformatika szoftvereinek fejlődésével és felhasználói körének egyre szélesebb körű bővülésével a vizualizációs célú rácsmodell-alkalmazásokon túl a gridalapú elemzések fokozottabb terjedésére is számíthatunk. Ez a módszercsalád bár nem teljesen új keletű, előnyei közül több igazából mostanában, a szoftverkörnyezet lehetőségeinek nagyfokú szélesedése kapcsán válik nyilvánvalóvá. A grid számos előnye közül hármat érdemes talán ezúttal hangsúlyozni. Az egyik a hozzáférhető adatok köréhez kapcsolódik, ugyanis

egyre több statisztikai szervezet vagy adatszolgáltató tesz közzé, vagy épp a közeljövőben tervezi, hogy közzétesz gridtípusú adatokat, köztük a KSH is. Ezek egyértelműen a gridmetodika alapján dolgozhatók fel a legmegfelelőbb módon. A rácsmodellek egy másik, fentebb többször hangsúlyozott elvi előnye az egységes rácsméretből fakad. Jelentős pozitívum ugyanis, hogy a modell kiegyenlíti a kicsi és a nagy adminisztratív közigazgatási egységek térképi méretkülönbségeit, egységes területi arányokat kialakítva az ábrákon. Persze hátrányként említhető egyúttal az is, hogy az adminisztratív beosztásoktól elrugaszkodó ábrázolások értelmezése és interpretációja ezzel valamelyest nehezebbé válik, így célszerű a megfelelő egyensúlyt megtalálni a rácsból kialakítása, a felbontás aggregáltsága és az értelmezhetőség lehetőségei között. Harmadik előnyként végül a rácsmodellek pszeudoraszter tulajdonságait érdemes kiemelni, azaz azt, hogy az alapvetően vektoros grid-állományokon raszteres műveleteket is végre lehet hajtani. Mindezzel lehetővé válik, hogy a raszteres térinformatika széles módszertani tárháza (lásd Szabó 2013) tovább bővítse a területi elemzők eszközkészletét.

IRODALOM

- Berry, J. K. (2013) *GIS Modeling: Applying Map Analysis Tools and Techniques* Basis Press, Fort Collins, Colorado, USA.
http://www.innovativegis.com/basis/BeyondMappingSeries/BeyondMapping_IV/Topic4/BM_IV_T4.htm#Section4
- Detrekői Ákos–Szabó György (2002): *Térinformatika* Nemzeti Tankönyvkiadó, Budapest.
- Dusek Tamás (2004): *Területi jövedelmi folyamatok Magyarországon* In: Barton G., Dormány G. (szerk.): *A magyar földrajz kurrens eredményei: II. Magyar Földrajzi Konferencia*, Szeged.
http://geography.hu/mfk2004/mfk2004/cikkek/dusek_tamas.pdf (letöltve: 2015. június)
- Eurostat (2006): *ESSnet GEOSTAT population-grid project* http://ec.europa.eu/eurostat/statistics-explained/index.php/Population_grids
- Eurostat (2013) Updated urban-rural typology: integration of NUTS 2010 and the latest population grid *Statistics in Focus* 16/2013. <http://ec.europa.eu/eurostat/en/web/products-statistics-in-focus/-/KS-SF-13-016> (letöltve: 2015. június)
- Fotheringham, A. S.–Brunsdon, C.–Charlton, M. E. (2000): *Quantitative geography: perspectives on spatial data analysis* Sage Publications. London, UK.
- Fotheringham, A. S.–Brunsdon, C.–Charlton, M. E. (2002): *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships* John Wiley & Sons Ltd. Chichester, UK.
- Funk, T. H.–Stiefer, P. S.–Chinkin, L. R. (2001): *Development of Gridded Allocation Factors for the State of Texas* Sanoma Technology Inc., Petaluma, USA.
<https://www.tceq.texas.gov/assets/public/implementation/air/am/contracts/reports/ei/DevelopmentOfGriddedSpatialAllocationFactors-update.pdf> (letöltve: 2015. június)
- G-Econ (2011): *Geographically based Economic data (G-Econ)* Yale University, New Haven, USA.
<http://gecon.yale.edu/>
- Hengl, T. (2006): Finding the right pixel size *Computers & Geosciences* 32 (9): 1283–1298.
- Nemes Nagy József–Jakobi Ákos–Németh Nándor (2001): A jövedelemegyenlőtlenségek térségi és település-szerkezeti összetevői *Statisztikai Szemle* 79 (10-11): 862–884.
- Németh Nándor–Kiss János Péter (2007): Megyéink és kistérségeink belső jövedelmi tagoltsága *Területi Statisztika* 47 (1): 20–45.
- Nordhaus, W.–Azam, Q.–Corderi, D.–Hood, K.–Victor, N. M.–Mohammed, M.–Miltner, A.–Weiss, J. (2006): *The G-Econ Database on Gridded Output: Methods and Data* Yale University, New Haven, USA.
- Pénzes János (2011): A jövedelmi szempontból elmaradott települések területi átrendeződése a rendszerváltozás után *Földrajzi Közlemények* 135 (1): 59–69.
- Pénzes János (2012): A területi jövedelemegyenlőtlenségek tendenciái és meghatározó tényezői Magyarország fejlett és elmaradott régióiban In: Nyári D (szerk.) *Kockázat – Konfliktus – Kihívás* pp. 686–699.

- A VI. Magyar Földrajzi Konferencia, a MERIEXWA nyitókonferencia és a Geográfus Doktoranduszok Országos Konferenciájának Tanulmánykötete, Szeged.
- SEDAC (2000): *U.S. Census Grids Socioeconomic Data and Applications* Center, Columbia University, New York, USA. <http://sedac.ciesin.columbia.edu/data/collection/usgrid>
- Szabó Gergely (2013): *Raszter alapú térinformatikai rendszerek* Debreceni Egyetem Természetföldrajzi és Geoinformatikai Tanszék, Debrecen.
- Thomas, R. W. (1977) *An introduction to quadrat analysis. Concepts and Techniques in Modern Geography* No. 12, University of East Anglia, Norwich, UK.
- Tomlin, C. D. (1990): *Geographic information systems and cartographic modelling* Prentice-Hall, Englewood Cliffs, New Jersey, USA.

Kulcsszavak: grid, rácsmodell, pszeudoraszter módszerek, város-falu különbségek, jövedelem, térinformatika.

Resume

Typical spatial socio-economic analyses are applying vector data related to polygons of administrative units. Raster data are, however, make it possible to overcome on modifiable areal unit problem of spatial administrative divisions, since the biased image of maps with large or small polygons are replaced by equal size grids. These raster-like spatial units are although managed in vector environment, but it is also possible to perform traditional techniques of raster-analytics on them. This paper introduces some possibilities to apply raster methodologies on such social grid data.